Biomedical Optics EXPRESS

Hand acupuncture point localization method based on a dual-attention mechanism and cascade network model

HAO WANG,[1] LI LIU,[1,2,*] YING WANG,[1,2] AND SENHAO DU[1]

[1] *Faculty of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China*
[2] *Zhejiang Provincial Key Laboratory of Part Rolling Technology, Ningbo 315211, China*
[*] *liuli@nbu.edu.cn*

Abstract: Deep learning techniques have, to a certain extent, solved the problem of overreliance on clinical experience for traditional acupoint localization, but the accuracy and repetition rate of its localization still need to be improved. This paper proposes a hand acupoint localization method based on the dual-attention mechanism and cascade network model. First, by superimposing the dual-attention mechanism SE and CA in the YOLOv5 model and calculating the prior box size using K-means++ to optimize the hand location, we cascade the heatmap regression algorithm with HRNet as the backbone network to detect 21 predefined key points on the hand. Finally, "MF-cun" is combined to complete the acupoint localization. The FPS value is 35 and the average offset error value is 0.0269, which is much lower than the error threshold through dataset validation and real scene testing. The results show that this method can reduce the offset error value by more than 40% while ensuring real-time performance and can combat complex scenes such as unequal lighting, occlusion, and skin color interference.

1. Introduction

In international sporting events, athletes often suffer from overtraining or physical overload causing excessive stress on the muscles and nervous system, resulting in injury or an inability to continue high-intensity competition [1]. In Chinese medical theory [2], acupuncture points are considered to be special points in the body that can be stimulated to regulate the overall balance of the body and can help relieve pain and inflammation caused by overexertion of muscles and joints [3]. Traditional acupuncture point positioning relies excessively on clinical experience, resulting in "no expertise" or "no professional equipment" to find the correct point and miss the best treatment time, thus reducing the therapeutic effect.

Applying deep learning techniques in medicine has become widespread [4], bringing new opportunities and challenges to medical research and clinical applications. Deep learning techniques can aid in image analysis, disease diagnosis, genomics, and drug discovery. For example, Lai et al. [5] proposed a YOLOv5 × 6 model for rapidly detecting surgical gauze, allowing for real-time gauze tracking in laparoscopic surgery and assisting surgeons in recalling missing gauze positions. Weng et al. [6] introduced the WSYOLO model for recognizing dental imprints and clefts and their corresponding locations, helping doctors assess patient conditions. Li et al. [7] developed a hybrid deep learning network algorithm that uses Fast R-CNN to extract the tongue region, further calibrates and segments the region with VGG, and finally, employs GoogLeNet to judge a person's physical condition based on tongue images. Ragodos et al. [8] utilized convolutional neural networks and transfer learning to classify dental abnormalities and conducted corresponding evaluations.

Furthermore, numerous researchers have found that acupoint localization, which used to be a challenging problem with traditional methods, is no longer limited. By introducing deep learning techniques, acupoint localization can not only improve accuracy but also make acupoint

#501663
Journal © 2023

https://doi.org/10.1364/BOE.501663
Received 25 Jul 2023; revised 7 Oct 2023; accepted 15 Oct 2023; published 26 Oct 2023

stimulation more intelligent and personalized, contributing positively to patients' rehabilitation treatments. Currently, deep learning-based acupoint localization methods can be divided into two categories. The first category is the direct method, which involves creating datasets with acupoints as data labels based on supervised learning principles. By iteratively training, testing, and validating the datasets, the acupoint coordinates are mapped onto the target objects. For example, Sun et al. [9] trained a homemade acupoint dataset to locate acupoints on the forearm and proposed an offset error evaluation method for assessing acupoint localization accuracy. Lan et al. [10] used a 3DMM face model and trained it with predefined acupoint labels to accomplish acupoint localization for different faces. The second category is the indirect method, which combines regression models to obtain human key points/skeletal points. By establishing a connection between the acupoint coordinates defined by the WHO (World Health Organization) [11] and the skeletal points, the acupoint positions with detection regions are calculated using spatial geometric coordinates and mapped onto the target objects. For example, Masood et al. [12] proposed a three-dimensional prediction method for hand acupoints by integrating RGB-CNN and depth-CNN. They used Mediapipe technology to obtain key points on the hand and transformed them into acupoints using the "cun" measurement from traditional Chinese medicine theory, achieving hand acupoint localization. Zhang et al. [13] utilized face-mesh and hair-segmentation models for face and hair segmentation and determined the position of ear acupoints. By calculating the "cun" measurement through face alignment, they obtained a set of points on the unconstrained ears, ultimately identifying ear acupoints. Chan et al. [14] proposed an SSD MobileNet deep learning network approach to detect body parts and calculated multiple acupoints on the arm by incorporating the "cun" measurement.

Each of the above acupoint location methods has its own advantages in the process of use. The direct method, by virtue of its own characteristics, can find acupuncture point locations faster, but because the human body varies in size and skeleton, it is not possible to accurately locate acupuncture points or even invalidate them for user groups not in the database, and its repeatability rate is low. Additionally, similar to smaller areas such as the hands and face where there are adjacent acupuncture points in close proximity, the direct method has difficulty mapping acupuncture points. The indirect method has strong scalability and can evolve all acupoint through individual key points, but currently, the overall accuracy is low and causes the position of acupoint to shift or even fail in the presence of unequal lighting, occlusion, and complex scenes.

To address the problem of hand acupoint localization based on the indirect method, this paper proposes a cascade network model with a superimposed dual-attention mechanism to ensure real-time performance while improving acupoint localization accuracy. Through experiments, the performance of the model in this paper is tested, and the feasibility of the method is verified by the offset error and other indicators.

2. Method

2.1. System architecture

To reduce the offset error, optimize the output of acupuncture points, and avoid the error caused by the assimilation of hand feature points by the external environment, this paper extracts feature points efficiently by superimposing a dual-attention mechanism, discards the single network regression algorithm and adopts a cascaded network model approach to achieve box selection before finding points. The main architecture in this paper is the cascaded SC-YOLOv5 algorithm and the heatmap regression algorithm with HRNet as the backbone network to detect 21 predefined key points on the hand in real scenes. In the detection process, the real-time images captured by the camera are fed into the SC-YOLOv5 algorithm network for detection, and the hand prediction bounding boxes are generated after several convolution and pooling operations. The predicted bounding box is used as the input for the next stage of key point regression, and the corresponding heatmap is generated by each branch in the HRNet network. The different resolution heatmap outputs from HRNet are fused to obtain a unified, high-precision heatmap, and the 21 predefined key point locations of the hand are marked in the image. To capture the hand acupuncture points, this paper combines the relationship between each acupuncture point and hand key points and "MF-cun", converts the 21 predefined key points to 2D coordinates using OpenCV, calculates the coordinates of hand acupuncture points, and maps acupuncture points. The system network architecture of this paper is shown in Fig. 1.

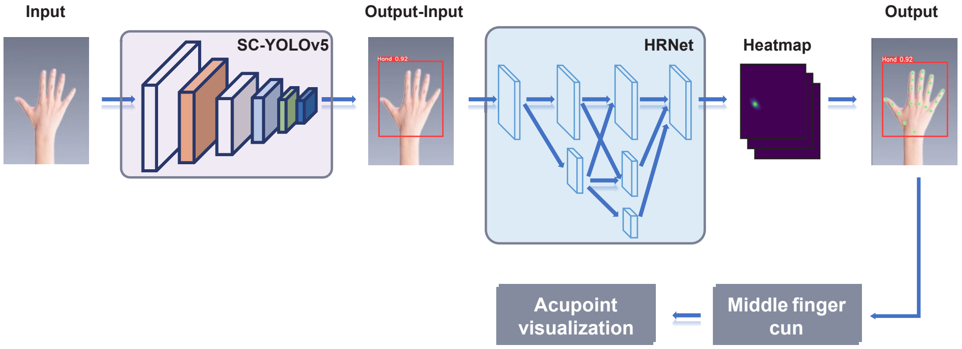

Fig. 1. System framework.

2.2. YOLOv5 network

YOLOv5 (you only look once version 5) [15] is a lightweight and novel network that treats object detection as a regression problem for classification and localization, enabling real-time detection and object classification in images. Compared to other versions, the YOLOv5 algorithm shows excellent performance in terms of detection accuracy and speed. Its structure can be divided into three parts: backbone, neck, and head [16]. Backbone CSPDarkNet is the base network, and its CSP (cross-stage partial connection) network structure ensures information flow while reducing computational complexity and improving detection speed. The neck module employs an FPN (feature pyramid network) [17] and PAN (path aggregation network) [18] for feature extraction and fusion. FPN extracts high-dimensional semantic information through top-down feature fusion, while PAN aggregates information from different levels of features in a bottom-up manner to capture detailed information and enhance detection accuracy. The head module adopts an anchor-free-like design, consisting of a classification module, a regression module, and an SPP (spatial pyramid pooling) module. The classification and regression modules fully extract features from various feature layers, and the SPP module performs multiscale cascaded pooling on the feature maps to further enhance detection performance. Currently, there are four versions of YOLOv5, namely, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. The four versions have the same network structure, different parameters, and different performance and speed. In this paper, the algorithm needs to be real-time while improving the accuracy, so YOLOv5s, which has the fewest parameters and is relatively the fastest, is selected.

2.3. Attention mechanism

This section introduces two attention mechanisms: squeeze-and-excitation (SE) [19] and coordinate attention (CA) [20].

SE enables the network to adaptively focus on important feature channels and reduce the reliance on unimportant feature channels by learning the weights among channels, which helps to improve the feature representation and robustness of the network and achieve better performance in various vision tasks. SE involves both squeezing and excitation processes.

Biomedical Optics EXPRESS

In the squeezing process, global average pooling is applied to encode all spatial features on each channel into a global feature. The output feature map has dimensions of $1 \times 1 \times C$, and the information for each channel Z_c is determined by Eq. (1).

$$Z_{\mathrm{c}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{\mathrm{j}=1}^{W} u_c(i, j) \tag{1}$$

where H and W are the spatial dimensions and u_c is the feature mapping obtained from the convolution of each channel.

During the excitation process, the relationship between different channels is obtained in the fully connected layer by the activation function sigmoid, and the importance of each channel is predicted to obtain the excitation weights. Finally, the excitation weights are multiplied with the original feature map in terms of elements for the recalibration of channel features.

CA provides the model's ability to perceive the location information in the input data for modeling spatial structure, improving spatial perception performance, handling scale changes and improving long-term dependence. Its structure consists of coordinate information embedding and coordinate attention generation.

In coordinate embedding, horizontal direction is encoded by applying a one-dimensional convolution operation along the vertical dimension of the feature tensor, resulting in the corresponding horizontal encoding feature. Similarly, vertical direction is encoded by applying a one-dimensional convolution operation along the horizontal dimension of the feature tensor, resulting in the corresponding vertical encoding feature. After embedding the coordinate information, the horizontal encoding feature and vertical encoding feature are used to generate an attention map that represents the target location information. This is achieved by performing an outer product operation between the horizontal encoding feature and the vertical encoding feature, resulting in an attention map. Finally, the generated attention map is elementwise multiplied with the input feature map to obtain a weighted feature map.

2.4. Optimization and improvement of YOLOv5

SE allows the network to automatically learn the weights for each channel, focusing more attention on channels with richer information. However, it only considers the interdependencies between channels and neglects spatial features. CA captures interchannel dependencies as well as direction-aware and position-aware information, thus compensating for the limitations of SE. Therefore, in this paper, adding a dual-attention mechanism (SE-CA) in the C3 module of the YOLOv5 model is proposed, and the improved C3 module is named “SC-C3”, as illustrated in Fig. 2(c).

The bottleneck in the original C3 module has three main components, as shown in Fig. 2(a). First, the initial feature map is passed through a 1×1 convolutional layer (CBS,Conv-BatchNorm-SiLU) to obtain F_A (feature map A). Then, F_A is passed through a 3×3 convolutional layer to obtain F_B, and F_B is added with the initial feature map to obtain the final output F_C.

In this paper, the dual-attention mechanism is superimposed after the 3×3 convolutional layer of bottleneck and named BottleNeck* with the structure shown in Fig. 2(b). First, the initial feature map is passed through a 1×1 convolution layer to obtain F_A and then a 3×3 convolution layer to obtain F_B. Second, F_B is passed to SE as input for global average pooling, converting the multichannel feature map into a vector of channels, mapping the output of the fully connected layer to weights between 0 and 1 using a sigmoid function, and multiplying the weights with F_B at the element level. F_B is passed to the CA as input, and the weights are determined by the sigmoid function as weights on the channel dimension and weighted summed with each channel of F_B to obtain F_E weighted by the attention of the CA. F_D and F_E are weighted and summed to obtain F_F. Finally, F_F is summed with the initial feature map to obtain the final output $F_{C'}$. The

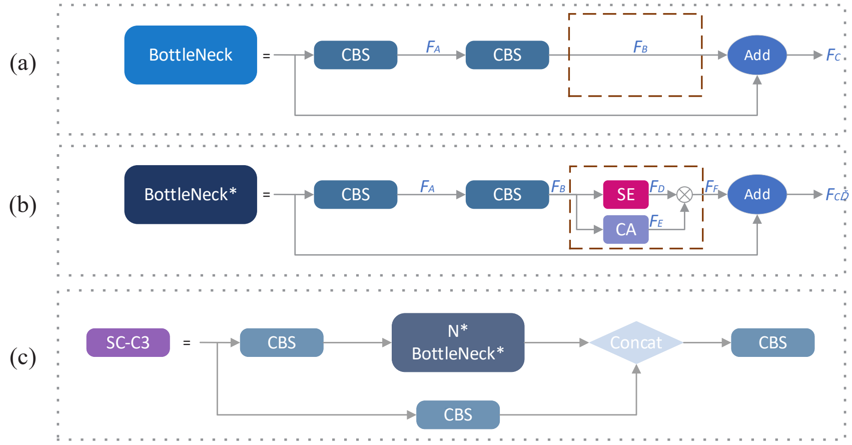

Fig. 2. (a) Original BottleNeck module, (b) improved BottleNeck* module, and (c) SC-C3 module.

weighted summation of F_F is determined by Eq. (2).

$$F_F = \alpha \times F_D + \beta \times F_E \tag{2}$$

where α and β are hyperparameters that control the attention weights of SE and CA.

Based on the SC-C3 module proposed above, inserted into the backbone of the YOLOv5 model, the improved model structure is shown in Table 1.

2.5. K-means++ optimizes priori box size

The model converges more easily when the size and scale of the a priori box are closer to the real bounding box. This is because the model can match the target object more accurately and reduce the prediction error by training the prior box parameters similar to the real bounding box. The prior box parameters of the original YOLOv5 model are calculated by matching the COCO dataset through the K-means algorithm [21], and its initial clustering centers are randomly selected, which are prone to fall into local minima, affecting the clustering effect of the bounding box size. K-means++ [22] clustering algorithm is an optimization algorithm based on K-means algorithm. Its main purpose is to improve the selection of initial points to make the anchor box size of the training dataset more appropriate and to improve the accuracy of the model in detecting objects. Therefore, we selected the K-means++ clustering algorithm as the priori box clustering method. The steps are as follows:

S1: Randomly select a sample from the dataset as the initial cluster center C_1.

S2: Calculate the minimum distance $D(x)$ between each sample and the existing cluster centers.

S3: Compute the probability P of each sample being selected as the next cluster center and choose the sample with the highest probability as the next generation cluster center, determined by Eq. (3) in the paper.

$$P = \frac{D(x)^2}{\sum_{i=1}^{n} D(x_i)^2} \tag{3}$$

S4: Repeat the work of S2 and S3 until 9 cluster centers are determined.

The prior box size parameters calculated according to the K-means++ and the original parameters by K-means are shown in Table 2.

Table 1. SC-YOLOv5 model structure

Num	From	Params	Module	Arguments
0	-1	3,520	Conv	[3, 31, 6, 2, 2]
1	-1	18,560	Conv	[31, 64, 3, 2]
2	-1	14,584	SCC3	[64, 64, 1]
3	-1	73,984	Conv	[64, 128, 3, 2]
4	-1	167,688	SCC3	[128, 128, 2]
5	-1	295,424	Conv	[128, 256, 3, 2]
6	-1	656,328	SCC3	[256, 256, 3]
7	-1	1,180,672	Conv	[256, 512, 3, 2]
8	-1	656,896	SPP	[512, 512, 5, 9, 13]
9	-1	526,336	SCC3	[512, 512, 1]
10	-1	131,584	Conv	[512, 256, 1, 1]
11	-1	0	Upsample	[None, 2,'nearest']
12	[-1, 6]	0	Concat	[1]
13	-1	361,984	C3	[512, 256, 1, False]
14	-1	33,024	Conv	[256, 128, 1, 1]
15	-1	0	Upsample	[None, 2,'nearest']
16	[-1, 4]	0	Concat	[1]
17	-1	90,880	C3	[256, 128, 1, False]
18	-1	147,712	Conv	[128, 128, 3, 2]
19	[-1, 14]	0	Concat	[1]
20	-1	296,448	C3	[256, 256, 1, False]
21	-1	590,336	Conv	[256, 256, 3, 2]
22	[-1, 10]	0	Concat	[1]
23	-1	1,182,720	C3	[512, 512, 1, False]

Table 2. Anchors parameter

Feature Map	Size(K-means++)			Size(K-means)		
Small	(7,10)	(19,12)	(25,30)	(10,13)	(16,30)	(33,23)
Middle	(15,20)	(40,26)	(53,62)	(30,61)	(62,45)	(59,119)
Large	(129,85)	(96,175)	(143,287)	(116,90)	(156,198)	(373,326)

2.6. Hand keypoint – heatmap regression

Heatmap regression, typically performed using convolutional neural networks (CNNs), is trained and predicted in specific network architectures chosen and optimized based on task requirements and dataset characteristics. Commonly used heatmap regression networks include HGNet (the hourglass network) [23], CPN (network with cross-pose) [24], OpenPose [25], AlphaPose [26], and HRNet [27]. In this paper, following the method described in Ref. [28], HRNet is chosen as the backbone network for heatmap regression and trained on a dataset of hand images to obtain 21 predefined key points [29], as shown in Fig. 3.

HRNet is a high-resolution network structure that employs a multibranch parallel approach, where each branch is responsible for extracting semantic features at different scales and levels. During heatmap regression, the position of each key point can be viewed as a Gaussian distribution on the heatmap. These distributions are stacked together with the feature maps, and accurate

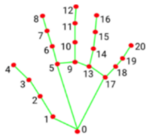

Fig. 3. Twenty-one predefined key points of the hand.

key point localization results are obtained through multilevel feature fusion. Compared to other network architectures, HRNet offers higher computational efficiency and improved accuracy. It effectively integrates information from different scales in the image, leading to further improvements in accuracy.

Table 3. Acupoints on the palm of the hand

Num	Acupoint	Acupoint location
1	Xiao chang dian	The midpoint of the transverse line between the first and second phalanges of the index finger
2	Da chang dian	The midpoint of the transverse line between the second and third phalanges of the index finger
3	Xin dian	The middle point of the transverse stripes between the second and third phalanges
4	Zhong chong	The midpoint of the fingertip, approximately 0.1 cun from the free edge of the nail
5	San jiao	The middle point of the transverse stripes between the first and second phalanges of the middle finger
6	Fei dian	The middle point of the transverse stripes between the second and third phalanges of the ring finger
7	Gan dian	The middle point between the first and second phalanges of the ring finger
8	Shen dian	The middle point of the transverse stripe of the second knuckle of the little finger
9	Shao fu	Between the fourth and fifth metacarpals, when making a fist, when the tip of the little finger
10	Lao gong	The second and third metacarpal bones are biased to the third metacarpal bones, and the middle fingertips are clenched when the fingers are flexed

Table 4. Acupoints on the back of the hand

Num	Acupoint	Acupoint location
1	Da gu kong	The midpoint of the transverse line of the interphalangeal joint of the thumb
2	San jian	The second metacarpophalangeal joint of the hand
3	Shang yang	The distal radial side of the index finger is 0.1 cun from the nail angle
4	He gu	Between the first and second metacarpals, the midpoint of the radial side of the second metacarpal
5	Zhong kui	The midpoint of the transverse line of the second joint
6	Ye meng	The fourth and fifth fingers, refers to the rear of the webbed edge of the red and white flesh
7	Guan chong	The ulnar end of the ring finger of the hand, 0.1 cun from the nail angle
8	Shao chong	0.1 cun posterior to the corner of nail on little finger’s radial side
9	Shao ze	Distal ulnar side of little finger, 0.1 cun above the corner side of nail root
10	Yang chi	On the dorsal wrist stripe, in the ulnar depression of the extensor tendon

2.7. Definition of hand acupuncture points

There are four main methods for acupoint localization in the human body, including surface anatomy landmarks, bone proportional measurement, finger cun measurement, and simplified methods (empirical methods) [11]. To improve the accuracy and robustness of hand acupoints localization while ensuring real-time detection, this paper selects the "MF-cun" (middle finger cun) method as the basis for acupoint localization. The "MF-cun" method considers the distance between the transverse creases at the ends of the inner side of the middle finger's middle joint when the finger is flexed as 1 "cun", which can be used for direct measurement of acupoints on the limbs and transverse measurement of acupoints on the back. In this paper, the distance between numbered 10 and numbered 11 in Fig. 3 is defined as the "cun" distance.

In this paper, the cascade deep learning algorithm is combined with "MF-cun" for accurate localization of hand acupoints. In Tables 3 and 4, the names and corresponding location information of 20 acupuncture points in the palm and dorsal region of the hand are described and marked with different color dots, as shown in Fig. 4.

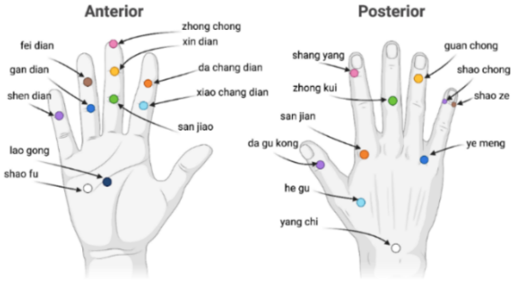

Fig. 4. Acupoints on the palm and back of the hand.

3. Results

3.1. Experimental environment

The hardware environment for this study consists of an Intel Core i7-10700 CPU @ 2.90 GHz, 16 GB of RAM, and Windows 10 64-bit operating system. The study utilizes the PyTorch 3.7.11 framework, CUDA version 10.2, and cuDNN version 7.6.5. The initial learning rate is set to 0.001, and the training process runs for 300 epochs.

3.2. Experimental dataset building

To validate the applicability of the proposed SC-YOLOv5 network model in real-world scenarios and the effectiveness of the overall approach, this study requires the creation of a corresponding dataset for training and testing. The dataset consists of HaGRID [30], 11 K Hands [31], and the Large-scale Multiview 3D Hand Pose Dataset [32]. Based on different hand gestures, different illumination levels and different skin tones as filtering conditions, HaGRID adopts 3,500, 11 K Hands adopts 1,500 as the target detection dataset, and the Large-scale Multiview 3D Hand Pose Dataset adopts 2,000 as the key point detection dataset (the dataset has 21 keypoints of the hand). A sample of the dataset is shown in Fig. 5.

To ensure accurate positioning, the data annotation software LabelMe is employed to precisely select and annotate the specific hand positions within the datasets. The annotated datasets are then divided into training and validation sets in an 8:2 ratio, facilitating subsequent evaluation and testing processes.

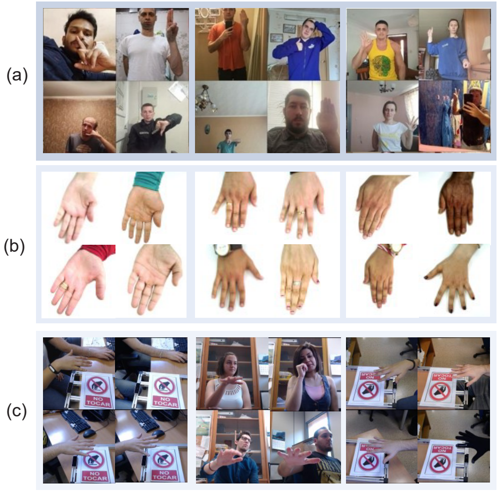

Fig. 5. Captured dataset: (a) Sample image from the HaGRID dataset, (b) Sample image from the 11 K Hands dataset, and (c) Sample image from the Large-scale Multiview 3D Hand Pose Dataset.

3.3. Visualization of hand acupoints

A random selection of individuals without hand disorders, aged between 20 and 30, was used to perform the detection of hand acupoints. To accurately evaluate the detection performance of the model and validate its robustness in real-world scenarios, this study conducted tests in different scenarios, including variations in lighting conditions, complex backgrounds, skin color interference, and occlusions. The results are shown in Fig. 6, showing from left to right the dorsal hand points, the reverse process points, and the palm points. The palm and back of the hand regions each show 10 different acupuncture points. The experiment proves that the algorithm in this paper has excellent robustness in all complex scenes and has certain antiocclusion capability, which can accurately return the acupoint on the palm and back of the hand to complete acupoint localization.

3.4. Evaluation parameter

To further validate the effectiveness of the proposed improvements to YOLOv5 and the efficacy of the method for hand acupoints localization, this section introduces quantitative evaluation metrics. For the SC-YOLOv5 algorithm, the following performance metrics are primarily utilized: precision (P), recall (R), F1 score, average precision (AP), mean average precision (mAP) and frames per second (FPS). These selected metrics are widely used in optical recognition tasks to evaluate detection accuracy. The calculation for the aforementioned performance metrics is presented in Equations (4-9).

$$\text{Precision}(P) = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4}$$

$$\text{Recall}(R) = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{6}$$

$$\text{Average Precision}(AP) = \int_0^1 P(r)dr \tag{7}$$

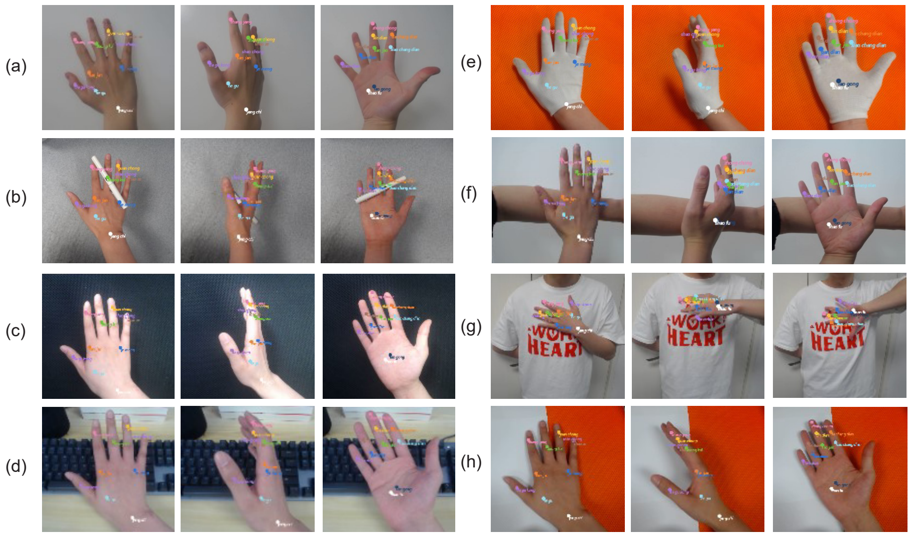

Fig. 6. Hand gesture detection results in real-world scenarios: (a) Normal light, (b) presence of occlusion, (c) bright light, (d) cluttered background, (e) wearing gloves, (f) overlapping with arm, (g) long-shot, and (h) different color backgrounds.

$$\text{Mean Average Precision}(mAP) = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{8}$$

$$\text{Frames Per Second}(FPS) = \frac{1}{preprocess + inference + NMS} \tag{9}$$

where true positive is the number of positive samples detected as positive, false positive is the number of negative samples detected as positive, false negative is the number of negative samples detected as negative, preprocess is the image preprocessing time, inference is the inference time, and NMS is the non-maximum suppression processing time.

To validate the effectiveness of the cascade network model combined with the "MF-cun" method for hand acupoint localization, it is necessary to ensure that the detected acupoint errors do not exceed their respective thresholds. However, due to the presence of varying scales in the images, direct measurement of errors using Euclidean distance is not feasible. Therefore, normalization is needed. In this study, the offset error is evaluated based on the approach described in [9], and its numerical value is determined using Eq. (10).

$$\text{offset error} = \frac{||p_i - \hat{p}_i||_2}{d} \tag{10}$$

In the equation, p_i represents the true acupoint label, $\widehat{p_i}$ represents the predicted acupoint label, and d is the scale normalization factor. In this study, the distance of d is determined as the distance between the labeled points 0 and 12, as indicated in Fig. 3.

3.5. Evaluation results of the SC-YOLOv5 algorithm

During training, based on the dataset partitioning, the loss variations are shown in Fig. 7. The result value for train/boxloss is approximately 0.0097, and for train/objloss, it is approximately 0.0055. The values for val/boxloss and val/objloss are approximately 0.018 and 0.0076, respectively.

According to the statistical data in Table 5, the SC-YOLOv5 model, improved through the approach described in Table 1, performs well in the evaluation metrics. The algorithm achieves

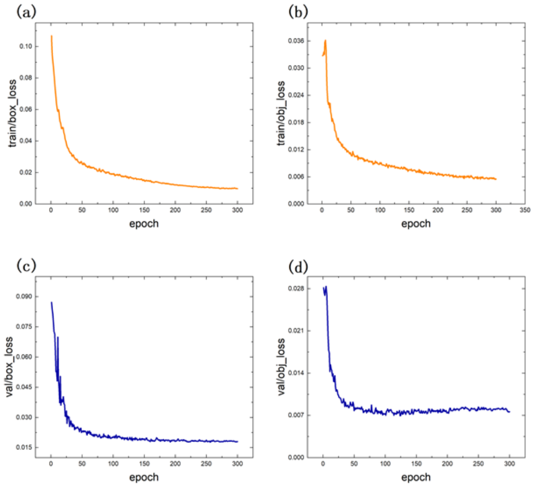

Fig. 7. Loss variation: (a) boundary loss on the training set, (b) average object detection loss on the training set, (c) boundary loss on the validation set, (d) average object detection loss on the validation set.

an accuracy of 97.75% on the validation set, showing an improvement of 1.97% compared to the original YOLOv5 model. It also outperforms SE-YOLOv5 (with only the SE attention mechanism) and CA-YOLOv5 (with only the CA attention mechanism). In addition, we also conducted a comparison of K-YOLOv5 (K-means++ clustering algorithm with improved a priori box parameters) with the original YOLOv5, which shows a slight improvement in all metrics, e.g., Precision is improved by 0.05% and Recall is improved by 0.11%. The feasibility of using K-means++ clustering algorithm in Section 2.5 is further demonstrated.

Table 5. The experimental effects of different networks

Model	AP/%	Precision/%	Recall/%	F1	mAP@0.5/%	FPS
YOLOv5s	91.92	95.78	93.32	94.64	96.29	76.4
K-YOLOv5	91.98	95.83	93.44	94.69	96.31	75.7
SE-YOLOv5	92.47	95.86	93.49	94.66	96.33	51.5
CA-YOLOv5	92.53	96.01	93.67	94.82	96.35	47.5
YOLOv8n	89.26	93.88	87.64	90.65	91.09	10.5
YOLOv8s	94.56	98.33	92.78	95.47	96.99	5.4
SC-YOLOv5	93.68	97.75	95.69	96.71	97.15	39.33

It is worth noting that SC-YOLOv5 is on par with YOLOv8 in terms of accuracy, but SC-YOLOv5 (39.99 fps) outperforms YOLOv8 (5.4 fps) in terms of real time, which suggests that the improved model in this paper can accurately detect the hand position while guaranteeing real-time performance.

3.6. Results of acupoint detection and evaluation

In this study, during the calculations, the normalization factor “d” is determined to be 18 cm. Sun et al. [9] mentioned that the finger contact area during pressing is 2 cm, and Lin et al. [33] suggested maintaining a safe distance of 3 cm from acupoints during moxibustion. To accommodate these conditions, this study takes $p_i - \widehat{p}_{i2}$ as 2 cm and calculates the offset error threshold as 0.111 using Eq. (10).

Based on the proposed method in this study for detecting hand acupoints, 200 sets of coordinate data are randomly selected as predicted acupoint values, while the true acupoint values are manually labeled values from Table 4 and Table 5. The average offset error (AOE) for each hand acupoint is calculated by combining Eq. (10). The results are shown in Table 6. From Table 6, it can be observed that the overall range of offset errors using this method is between 0.01 and 0.06, with an AOE of 0.0269. The relatively higher AOE for acupoints such as “lao gong”, “shao fu”, “shao ze” and “shao chong” are due to the inherent deviations in the “MF-cun” calculation process, but the overall fluctuations are not particularly significant.

Table 6. Average offset error value of acupoint

Acupoint	AOE	Acupoint	AOE
xiao chang dian	0.0158	Da gu kong	0.0108
da chang dian	0.0135	San jian	0.0323
Xin dian	0.0119	Shang yang	0.0393
Zhong chong	0.0187	He gu	0.0325
San jiao	0.0276	Zhong kui	0.0154
fei dian	0.0281	Ye meng	0.0329
gan dian	0.0152	Guan chong	0.0122
Shen dian	0.0199	Shao chong	0.0534
Shao fu	0.0430	Shao ze	0.0586
Lao gong	0.0489	Yang chi	0.0198

In addition, Table 7 provides a comparison of acupoint localization based on different algorithms, focusing on detection areas such as the face, forearm, ear, and hand. References [9,10] employ direct methods for acupoint localization, while Refs. [12–14] and this study utilize indirect methods. Among them, Refs. [10,13] mention the ability to resist partial occlusion during acupoint detection. The localization method described in Ref. [10] is applied to mobile devices, resulting in slightly lower FPS.

Table 7. Compare the different methods of locating acupoints

References	Surveyed area	Gordian technique	Resistant occlusion	AOE	FPS	Limitation
Sun et al. [9]	Forearm	VGG-19 and acupoint database	NO	–	–	Requires presetting acupoint locations
Lan et al. [10]	face	3DMM and Landmark detection	YES	–	7.5	Requires presetting acupoint locations
Masood et al. [12]	hand	Feature extraction and landmark detection based on CNN	NO	0.0532	–	Hand needs to be stretched
Zhang et al. [13]	Ear	Ear recognition framework and B-cun	YES	0.039	30	The acquisition of acupoint locations relies on facial alignment
Chan et al. [14]	Forearm	SSD MobileNet and Mesh generation	NO	0.0585	–	The acquisition of acupoint locations relies on grid generation
Present study	hand	Based on SC-YOLOv5 and HRNet dual network models and “MF-cun”	YES	0.0269	35	The tested hands need to meet no mutilation

Biomedical Optics EXPRESS

4. Discussion and summary

The proposed method in this paper utilizes a cascade deep learning network model with a stacked dual-attention mechanism to accurately regress acupoint locations. According to the data in Table 7, compared to other methods, the AOE of the proposed method is reduced by more than 40%. When calculating the offset error, a normalization factor is used, which depends on the maximum size of the detection area. According to Eq. (10), under other unchanged conditions, the smaller the normalization factor is, the larger the resulting offset error. In this study, the normalization factor is similar to that in Ref. [13] but smaller than the value in Ref. [14]. Therefore, in terms of accuracy, the proposed method has a significant advantage.

Furthermore, as shown in Fig. 6, the proposed method is robust to uneven lighting conditions, skin tone interference, occlusions, and other complex backgrounds. Compared to Refs. [9,10], this paper employs an indirect method for acupoint localization, eliminating the need to establish a large dataset or predefine acupoint locations. Compared to Refs. [13,14], the proposed method incorporates its own “MF-cun” calculation for acupoint localization and does not require conditions such as reference point alignment. In summary, the proposed method exhibits significant advantages in terms of robustness.

In this paper, we propose a hand acupoint localization method based on a dual-attention mechanism and a cascaded deep learning network model to achieve accurate localization of hand acupoint points under complex backgrounds such as unequal illumination, presence of occlusion, and skin color interference.

In the detection, firstly, SC-YOLOv5 is used for accurate bounding box selection of hand locations, which greatly enhances the detection capability of hand feature points and effectively mitigates the interference of complex backgrounds. Then, HRNet based heatmap regression method was used to accurately regress the hand key points, and then combined with “MF-cun” to obtain the hand acupuncture points by OpenCV calculation. Experiments show that the average offset error of the detected acupoints is 0.0269, which is more than 40% lower than other methods.

In the future, our plans will focus on acupoint localization in user groups with limb defects or specific medical conditions. For such user groups, due to the lack of readily available databases and the inability of the created datasets to cover certain unique cases, it may be challenging to accurately locate acupoints. One potential direction to address this issue is to explore techniques such as compensatory restoration to restore the original condition of the user’s limb before conducting acupoint localization.

Funding. National Natural Science Foundation of China (51875294); Zhejiang Province Public Welfare Technology Application Research Project (LGF21H180004).

Disclosures. The authors declare no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request. The public datasets are available in Refs. [30,31].

References

1. A. E. Saw, L. C. Main, and P. B. Gastin, “Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review,” Br. J. Sports Med. **50**(5), 281–291 (2016).
2. Y. H. Tsai, S. Y. Wu, W. L. Hu, Y. R. Lai, Y. Tsao, K. T. Yen, C. H. Lin, and C. A. Kuo, “Immediate effect of non-invasive auricular acupoint stimulation on the performance and meridian activities of archery athletes: A protocol for randomized controlled trial,” Medicine **100**(8), e24753 (2021).
3. F. Li, T. He, Q. Xu, L.-T. Lin, H. Li, Y. Liu, G.-X. Shi, and C.-Z. Liu, “What is the Acupoint? A preliminary review of Acupoints,” Pain Med. **16**(10), 1905–1915 (2015).
4. R. Daneshjou, B. He, D. Ouyang, and J. Y. Zou, “How to evaluate deep learning for cancer diagnostics - factors and recommendations,” Biochim. Biophys. Acta, Rev. Cancer **1875**(2), 188515 (2021).
5. S. L. Lai, C. S. Chen, B. R. Lin, and R. F. Chang, “Intraoperative Detection of Surgical Gauze Using Deep Convolutional Neural Network,” Ann. Biomed. Eng. **51**(2), 352–362 (2023).
6. H. Weng, L. Li, H. Lei, Z. Luo, C. Li, and S. Li, “A weakly supervised tooth-mark and crack detection method in tongue image,” Concurrency and Computation-Practice & Experience **33**(16), e6262 (2021).

7. H. H. Li, G. H. Wen, and H. B. Zeng, "Natural tongue physique identification using hybrid deep learning methods," Multimed. Tools Appl. **78**(6), 6847–6868 (2019).
8. R. Ragodos, T. Wang, C. Padilla, J. T. Hecht, F. A. Poletta, I. M. Orioli, C. J. Buxó, A. Butali, C. Valencia-Ramirez, C. R. Muñeton, G. L. Wehby, S. M. Weinberg, M. L. Marazita, L. M. M. Uribe, and B. J. Howe, "Dental anomaly detection using intraoral photos via deep learning," Sci. Rep. **12**(1), 11577 (2022).
9. L Sun, S Sun, Y Fu, and X Zhao, "Acupoint detection based on deep convolutional neural network," in *2020 39th Chinese control conference (CCC)* (2020), pp. 7418–7422.
10. K. C. Lan, M. C. Hu, Y. Z. Chen, and J. X. Zhang, "The Application of 3D Morphable Model (3DMM) for Real-Time Visualization of Acupoints on a Smartphone," IEEE Sens. J. **21**(3), 3289–3300 (2021).
11. S. Lim, "WHO Standard Acupuncture Point Locations," Evidence-Based Complementary and Alternative Medicine **7**(2), 167–168 (2010).
12. D. Masood and J. Qi, "3D Localization of Hand Acupoints Using Hand Geometry and Landmark Points Based on RGB-D CNN Fusion," Ann. Biomed. Eng. **50**(9), 1103–1115 (2022).
13. M. H. Zhang, J. P. Schulze, and D. Zhang, "E-faceatlasAR: extend atlas of facial acupuncture points with auricular maps in augmented reality for self-acupressure," Virtual Reality **26**(4), 1763–1776 (2022).
14. T.W. Chan, C. Zhang, W.H. Ip, and A.W Choy, "A Combined Deep Learning and Anatomical Inch Measurement Approach to Robotic Acupuncture Points Positioning," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Annual International Conference 2021*, 2597–2600 (2021).
15. U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs," Sensors **22**(2), 464 (2022).
16. Y. T. Wu, S. M. Tang, S. W. Zhang, and H. Ogai, "An Enhanced Feature Pyramid Object Detection Network for Autonomous Driving," Appl. Sci. **9**(20), 4363 (2019).
17. J. F. Hu, J. Sun, Z. Lin, J. H. Lai, W. Zeng, and W. S. Zheng, "APANet: Auto-Path Aggregation for Future Instance Segmentation Prediction," IEEE Trans. Pattern Anal. Mach. Intell. **44**(7), 1 (2021).
18. J. Qi, X. Liu, K. Liu, F. Xu, H. Guo, X. Tian, M. Li, Z. Bao, and Y. Li, "An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease," Computers and Electronics in Agriculture **194**, 106780 (2022).
19. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 7132–7141.
20. Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), pp. 13708–13717.
21. X. Wang and Y. Bai, "The global Minmax k-means algorithm," SpringerPlus **5**(1), 1665 (2016).
22. D Arthur and S Vassilvitskii, "K-means++ the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, (2007), pp. 1027–1035.
23. A Newell, K Yang, and J Deng, "Stacked hourglass networks for human pose estimation," in *Computer vision – ECCV 2016*. Cham: Springer International Publishing (2016), pp. 483–499.
24. Z Cao, T Simon, S-E Wei, and Y Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017), pp. 7291–7299.
25. Y. Konishi, K. Hattori, and M. Hashimoto, "Real-Time 6D Object Pose Estimation on CPU," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2019), pp. 3451–3458.
26. H-S Fang, J Li, H Tang, C Xu, H Zhu, Y Xiu, Y-L Li, and C Lu, "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time," in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022), pp. 7157–7173.
27. J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B Xiao, "Deep High-Resolution Representation Learning for Visual Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2021), pp. 3349–3364.
28. Y. W. Zhang, T. Zhu, H. S. Ning, and Z. Y. Liu, "Classroom student posture recognition based on an improved high-resolution network," J. Wireless Com. Network **2021**(1), 140 (2021).
29. J.-Y. Kim and K.-S. Kang, "Korean fingerprint recognition using hand landmark," in *J Korea Next Gener Comput Soc* (2022), pp. 81–91.
30. A Kapitanov, A Makhlyarchuk, and K Kvanchiani, "Hagrid-hand gesture recognition image dataset," arXiv, arXiv:220608219 (2022).
31. M. Afifi, "11 K hands: Gender recognition and biometric identification using a large dataset of hand images," Multimed. Tools Appl. **78**(15), 20835–20854 (2019).
32. F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Large-scale Multiview 3D Hand Pose Dataset," arXiv, arXiv:1707.03742 (2022).
33. L. M. Lin, S. F. Wang, R. P. Lee, B. G. Hsu, N. M. Tsai, and T. C. Peng, "Changes in skin surface temperature at an acupuncture point with moxibustion," Acupunct. Med. **31**(2), 195–201 (2013).